



Comment peut-on construire une ontologie personnelle à partir de textes ? Considérations linguistiques

Anne Condamines

► To cite this version:

Anne Condamines. Comment peut-on construire une ontologie personnelle à partir de textes ? Considérations linguistiques. Terminologie et Intelligence Artificielle 2009 (TIA'09), Nov 2009, Toulouse, France. halshs-01321023

HAL Id: halshs-01321023

<https://shs.hal.science/halshs-01321023>

Submitted on 13 Oct 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Anne Condamines
CNRS et Université de Toulouse
CLLE-ERSS (UMR 5263)
anne.condamines@univ-tlse2.fr
fax : 05 61 50 46 77

**Comment peut-on construire une ontologie personnelle à partir
de textes ? Considérations linguistiques**

Mots clés : Linguistique appliquée, Ontologie individuelle, Ontologie personnelle,
Ontologie locale, Subjectivité discursive, Terminologie textuelle,
Key words : Applied Linguistics, Individual Ontology, Linguistics subjectivity,
Local ontology, Personal ontology, Textual terminology.

Résumé : L'article s'intéresse d'un point de vue linguistique à la construction d'ontologies personnelles à partir de textes. Il vise à montrer les spécificités que la notion d'ontologie personnelle amène par rapport à celle d'ontologie régionale. Ces spécificités concernent la structure des ontologies elles-mêmes mais principalement, les caractéristiques des corpus nécessaires à ces constructions. Trois corpus (deux manuels et un entretien) sont comparés, selon certains traits linguistiques d'abord puis du point de vue des marqueurs de relations conceptuelles. Cette comparaison permet de conclure que la constitution d'ontologies personnelles à partir de textes, tout en n'étant pas impossible, pose des questions plus redoutables que la constitution d'ontologies régionales, en lien avec le mode de prise en charge de son discours par le locuteur individuel.

Abstract : This paper deals with personal ontologies construction from texts using a linguistic point of view. It aims to emphasize differences between personal ontology and local ontology. The points examined are variations concerning ontology structure itself but mainly, characteristics concerning corpora used for ontologies construction. Three corpora (two handbooks and a corpus of interviews) are compared, examining first some linguistics features and then conceptual relation patterns. The comparison leads to the conclusion that the construction of a personal ontology from texts, if not impossible, raises issues much more difficult than the construction of local ontologies, because it may be difficult to clarify how the individual writer takes into account his/her own position in his/her discourse.

Comment peut-on construire une ontologie personnelle à partir de textes ? Considérations linguistiques

Anne Condamines

CNRS et Université de Toulouse, CLLE-ERSS (UMR 5263)
anne.condamines@univ-tlse2.fr

Résumé : L'article s'intéresse d'un point de vue linguistique à la construction d'ontologies personnelles à partir de textes. Il vise à montrer les spécificités que la notion d'ontologie personnelle amène par rapport à celle d'ontologie régionale. Ces spécificités concernent la structure des ontologies elles-mêmes mais principalement, les caractéristiques des corpus nécessaires à ces constructions. Trois corpus (deux manuels et un entretien) sont comparés, selon certains traits linguistiques d'abord puis du point de vue des marqueurs de relations conceptuelles. Cette comparaison permet de conclure que la constitution d'ontologies personnelles à partir de textes, tout en n'étant pas impossible, pose des questions plus redoutables que la constitution d'ontologies régionales, en lien avec le mode de prise en charge de son discours par le locuteur individuel.

Mots-clés : Ontologie individuelle, Ontologie personnelle, Subjectivité discursive, Terminologie textuelle.

1 Introduction

La première partie du titre de cet article est à comprendre de deux façons. Il s'agit tout d'abord de se demander comment il est possible de parler d'ontologie personnelle alors que l'association de ces deux termes pourrait donner un effet d'oxymore. Il s'agit ensuite de voir comment une telle ontologie, si on l'admet possible, peut être construite à partir de textes. Cet article s'inscrit dans une perspective linguistique. En effet, si un certain nombre de travaux de terminologie textuelle se sont déjà penchés sur la constitution d'ontologies régionales à partir de textes (Condamines et Rebeyrolle, 2000), ce n'est pas le cas pour la constitution d'ontologies personnelles. Or, ce type d'ontologie présente des caractéristiques qui ont une influence non négligeable sur leur mode de constitution et la linguistique, dans cette tâche, peut avoir un intérêt pour éclairer les questionnements.

La deuxième partie consiste en une interrogation générale sur les ontologies personnelles. La troisième interroge les rapports entre textes et ontologie personnelle. La quatrième explore quelques phénomènes en corpus (plus particulièrement les marqueurs de relations conceptuelles) pour poser des jalons préliminaires concernant la possibilité de construire une ontologie personnelle à partir de textes.

2 Ontologies personnelles

Si l'on s'en tient à la définition du mot ontologie en philosophie (« ce qui est » ou « l'étant »), son association avec l'adjectif « personnelle » peut paraître assez surprenante. En effet, a priori, « ce qui est » n'est pas dépendant d'un individu.

Pourtant, d'une part, la notion d'ontologie personnelle existe en philosophie et d'autre part, on pourrait identifier la même contradiction avec l'expression « ontologie locale ou régionale » pourtant largement admise à présent en (Ingénierie des connaissances (désormais IC). Une recherche sur internet montre que le terme « personal ontology » est utilisé en philosophie. Il apparaît en particulier dans un ouvrage de Olson (2007). Mais si on se penche sur l'usage qui est fait de ce terme, on voit qu'il renvoie à des interrogations sur ce que signifie « être une personne », c'est-à-dire « an animal that is biologically human ». Il ne s'agit donc pas de s'interroger sur l'« ontologie personnelle » au sens d'« ontologie individuelle » même si l'auteur finit par questionner les rapports entre l'humanité biologique et la personnalité psychologique. En revanche, Gusdorf, lui aussi philosophe, utilise bien ce terme « d'ontologie personnelle » pour le renvoyer à une dimension psychologique, celle qui s'exprime dans la parole individuelle (Gusdorf, 1952). Cette dimension psychologique sera à interroger dans la constitution d'ontologie personnelle, la notion d'individu faisant nécessairement intervenir la dimension d'expérience personnelle. Pourtant, traditionnellement, dans la constitution d'ontologies en tout cas d'ontologies de domaine (ou encore ontologie locale ou régionale) c'est la dimension sociale qui est pertinente. Rappelons sans rentrer dans le détail que la notion d'ontologie locale ou régionale a dû être largement argumentée pour être acceptée par la communauté d'IC (Bachimont, 1995). Comme nous le soulignons, l'ajout de l'adjectif « locale » ou « régionale » introduit une rupture dans la nature de « ce qui est » et oblige à prendre en compte la notion de point de vue en fonction d'un domaine et/ou d'une application. « Ce qui est » ne serait donc pas unique mais dépendant d'un certain nombre d'éléments qui doivent être mis au jour pour que l'ontologie ait une pertinence.

La recherche sur internet de « personal ontology » en montre un usage relativement récent (une quinzaine d'années) et un usage encore plus récent de « ontologie personnelle ». On trouve d'ailleurs un certain nombre de concurrents à ces termes : *individual ontology*, *ontologie individuelle*, *local conceptualization*, *personnalisation d'ontologies* (ce dernier terme renvoyant d'ailleurs plus à une vision descendante, d'une ontologie existante à sa spécification pour des individus)... Dans tous les cas, l'idée sous-jacente est qu'il y a un sens à considérer une ontologie non pas, ou en tout cas pas d'emblée, dans ce qu'elle a de collectif mais dans ce qu'elle a d'individuel. Elle est ainsi définie par Jones (2005) :

“A personal ontology [by extension] is simply one persons ontology as opposed to a global, all encompassing ontology (e.g. the kind of grand unifying and restricting taxonomy-like structure that Sig is railing against). A personal ontology on the other hand is more relevant to the individual”.

Deux courants majeurs semblent utiliser cette notion d'ontologie personnelle. D'une part, la recherche d'information sur le web qui souhaite personnaliser le profil des

utilisateurs (Aimé et al, 2009), (Brusilovsky and Kobsa, 2007), d'autre part, tous les projets de construction collaborative qui souhaitent construire un consensus à partir des différents point de vue des individus concernés (Falquet et al., 2004), (Euzénat, 2009), (Jung, 2008). La constitution de ces ontologies personnelles fait souvent intervenir un « corpus fourni par l'utilisateur et supposé représentatif de son univers cognitif » (Aimé et al, 2009) ce qui inscrit cette problématique au moins en partie dans la problématique de la construction d'ontologies à partir de textes. L'objet de cet article est de poser les premières questions de ce que peut recouvrir une telle construction, d'un point de vue linguistique.

3 Que peut-il y avoir de personnel dans une ontologie construite à partir de textes ?

La notion d'individu dans « ontologie individuelle » ne va pas de soit. En effet, alors que les ontologies locales ou régionales font intervenir un point de vue collectif, sur un domaine, une activité, une application et sont donc soutenues par une forte dimension sociale, le point de vue d'un individu peut être nettement moins évident à cerner. En effet, la notion d'individualité fait entrer dans la problématique non plus seulement l'aspect sociologique mais aussi potentiellement l'aspect psychologique, en d'autres termes, l'expérience individuelle. Plus encore que pour les ontologies régionales, la question se pose de savoir ce qui va être retenu dans ces ontologies personnelles et comment on va passer d'un discours individuel (manifestant l'ontologie personnelle au sens de Gusdorf) à une représentation relationnelle de la connaissance. On va voir que, si les méthodes de construction désormais bien connues des ontologies à partir de textes peuvent être appliquées pour la construction d'ontologies personnelles, en particulier l'utilisation de marqueurs de relation, certaines difficultés apparaissent du fait de la nécessité de prendre en compte le point de vue d'un individu et plus particulièrement de le distinguer de celui d'un point de vue collectif.

3.1 Termes/Relations/marqueurs

La représentation sous forme de termes reliés par des relations constitue le dénominateur commun de tous les types d'ontologies (et de manière moins systématique, de terminologies). Les ontologies individuelles ne font pas exception à cette règle. On peut se demander quelle marge de manœuvre existe pour la construction d'ontologies personnelles, c'est-à-dire quel type de variation peut-être pris en compte. Construire une ontologie (ou une terminologie) à partir de textes revient à passer d'un système linéaire, syntaxique à un système volumique et iconique (nœuds et flèches), ce qui est toujours le résultat d'une interprétation (Condamines, 2007). Cette interprétation est faite le plus souvent par un autre que le rédacteur/locuteur qui met en œuvre un point de vue interprétatif déterminé *a priori* (objectif d'utilisation). On retrouve cette dimension interprétative dans la construction d'ontologies personnelles mais s'y surajoute un autre élément : rendre compte d'un

point de vue individuel. Si l'on admet avec les structuralistes que la représentation sous la forme d'un réseau de relations n'est autre qu'une façon de définir un terme (le sens de ce terme étant alors sa place dans le réseau), on peut considérer qu'une ontologie personnelle doit rendre compte d'une façon de « définir subjectivement ». Or, le principe même du fonctionnement langagier fait que les possibilités de définir subjectivement, c'est-à-dire d'une manière qui se démarque de ce qui est généralement admis dans la communauté des locuteurs, sont assez faibles.

En effet, un individu, qui est avant tout locuteur d'une langue qui, par essence, fonctionne sur un mode collectif, n'a que très peu de marge de manœuvre pour créer des termes ou des relations, c'est-à-dire se situer dans un processus néologique. S'il le fait, il doit, en principe, signaler cette création, soit en le disant explicitement (*j'appelle tel concept tel nom*), soit en le soulignant par une indication typographique (des guillemets par exemple) ou prosodique. Mais cette créativité ne peut être que marginale sinon le locuteur prendrait le risque d'être exclu de la communauté linguistique à laquelle malgré tout, il se rattache. Les marqueurs de relation, quant à eux, semblent à peu près exclus de toute créativité. En effet, de manière générale, ils font partie de processus métalinguistiques qui sont bien plus stables que les productions discursives et relèvent d'une compétence bien moins experte que celle qui concerne les termes. C'est en tout cas l'hypothèse qui sous-tend la constitution d'ontologies ou de terminologies à partir de textes. Notons d'ailleurs que la notion de « terminologie personnelle » (qui n'apparaît jamais sur internet) semblerait très problématique tant la terminologie, parce qu'elle est clairement inscrite dans un fonctionnement langagier, suppose l'adhésion à un fonctionnement collectif.

Si donc on accepte la notion d'ontologie personnelle, la seule façon de la prendre en compte concerne les choix de structuration, c'est-à-dire d'organisation des termes et des relations.

Cette différence dans les choix de structuration est en effet désormais acceptée et supportée par les outils d'IC. Elle est par ailleurs prise en compte en terminologie textuelle à travers la notion de multidimensionalité qui permet à un fils d'avoir plusieurs pères (Bowker, 1997) et, de manière plus linguistique, à travers la notion de point de vue, de sa mise au jour et de sa prise en compte dans des systèmes de termes (Condamines et Rebeyrolle, 1996).

En général, ces différences de sens (de structuration donc pour le point de vue structuraliste) sont repérées de deux manières : par des variations dans la mise en relation de termes en discours d'une part et/ou par des variations de distributions d'autre part. Nous ne traitons dans cet article que des variations explicites de mise en relation.

Le problème consistera donc en la prise en compte de l'aspect « subjectif » dans les textes, particulièrement en lien avec des marqueurs de relation.

3.2 Textes

Dans la construction de terminologies ou d'ontologies à partir de textes, la nature du corpus mis en œuvre joue un rôle déterminant. On sait ainsi qu'il existe des variations dans le fonctionnement des marqueurs de relation en fonction du genre

textuel. De fait, il semble établi que les corpus les plus riches en marqueurs de relations et donc les plus susceptibles d'être utilisés pour construire des ontologie/terminologies sont les corpus de type didactique comme les « manuels ». En effet, les auteurs de manuels utilisent fréquemment des contextes définitoires ou du moins des contextes riches en connaissance destinés à faire comprendre un domaine ou une pratique pour des non-experts. Ce genre de texte suppose de neutraliser la variation interindividuelle par l'hypothèse forte d'un « locuteur collectif », terme emprunté à la sociolinguistique. Le locuteur collectif est censé parler (ou écrire) en tant que représentant d'une communauté de locuteurs et donc de savoirs. C'est d'ailleurs pourquoi le discours attendu est censé être très neutre et donc dépourvu de marques de subjectivité.

Quid des textes censés être utilisés pour construire une ontologie personnelle ? On se trouve à nouveau devant une difficulté. En effet, à l'inverse des discours produits par de supposés « locuteurs collectifs », les corpus étudiés pour construire des ontologies personnelles devraient être riches en marques de subjectivité. Or, les corpus les plus riches en marqueurs de relations sont les corpus à visée didactique. Les notions de « didactique » et de « subjectivité » semblent assez contradictoires et il paraît très difficile d'identifier un genre textuel répondant à ces deux besoins. Il semble que, dans la plupart des cas de construction d'ontologie personnelle, on demande aux personnes concernées de produire des textes « représentatif de [leur] univers cognitif » ou bien qu'on utilise la transcription d'un entretien portant justement sur sa connaissance/compétence. Une différence majeure apparaît avec la constitution d'une ontologie régionale : le rédacteur du corpus et le valideur de l'ontologie (l'expert) sont obligatoirement la même personne.

Mais même dans les textes où la subjectivité est *a priori* très présente, les énoncés ne sont pas toujours pris en charge de manière nette par le locuteur. Nous reparlerons de ce problème dans le paragraphe 4.1.3, à propos de la prise en charge des marqueurs « définitoires ».

4 D'un corpus à une ontologie personnelle : étude de cas préliminaire

Cette partie vise à interroger les difficultés linguistiques auxquelles on peut se heurter lors de la constitution d'une ontologie personnelle à partir de textes.

L'étude comparative est menée sur trois corpus de taille assez proche : il s'agit de deux manuels (réputés riches en « contextes définitoires » ainsi que nous l'avons vu ci-dessus) et de la retranscription d'entretiens menés avec 5 experts en veille stratégique afin d'identifier leurs points de vue (projet Corinthe¹).

- Un manuel de mécanique, extrait de la base scitext de frantext. [Mécanique]
- Une partie (portant sur l'optoélectronique) d'un manuel rédigé par et pour le CNES. [Opto]
- La retranscription des 5 entretiens en lien avec le projet Corinthe [Corinthe].

¹ Ce projet, sous la responsabilité de P.-J. Charrel, a été financé par la Région Midi-Pyrénées

Une première étude va porter sur la comparaison « globale » des deux manuels et de l'entretien. La seconde étude concernera plus particulièrement les marqueurs de relation.

4.1.1 Caractéristiques générales

A priori, les trois corpus étudiés sont des corpus « spécialisés » (portant sur 3 domaines identifiables : mécanique, optoélectronique, veille stratégique), dont les locuteurs/rédacteurs sont des experts s'adressant à des novices ou des moins experts. Cette situation de « décalage dans les connaissances » (expert vs novice) est également présente dans tous les textes ainsi que la volonté de « définir » les connaissances.

Pourtant les quelques résultats qui suivent montrent que des différences importantes apparaissent entre d'une part les manuels et d'autre part le discours subjectif.

	Mécanique	Corinthe	Opto
Nbre de mots	105 700	92 000	109 500
Noms	27521 (26%)	14958 (16%)	35353 (32%)
Verbes	8559 (8%)	13295 (14%)	9740 (9%)
Nom de dét	893 (3,2%)	459 (3%)	1188 (3,3%)
Nom adj	3667 (13,3%)	1201 (8%)	4974 (14%)
Nom à dét	281(1%)	90 (0,6%)	293 (0,8%)
Adj N	1542 (5,6%)	1201(8%)	2223(6,3%)
Adjectifs	6870 (6%)	3814 (4%)	9124 (8%)
Adverbes	4239 (4%)	9562 (10%)	5354 (5%)

Table 1 : caractéristiques générales des corpus

Noms vs verbes

Il est traditionnellement admis que, dans les corpus spécialisés, les noms sont plus nombreux que les verbes (voir l'étude que nous avons menée dans Condamines et Bourigault, 1999). On considère en effet que, dans les textes de spécialité, les noms permettent de donner un effet d'objectivation que ne donnent pas les verbes et sont donc utilisés à leur place, sous une forme nominalisée (Rastier, 1995). On retrouve cette caractéristique dans les deux manuels mais pas dans l'entretien puisque là, il y a pratiquement autant de verbes que de noms. On peut donner deux explications pour ce phénomène : d'une part, l'entretien constitue justement un cas où il n'y a pas recherche d'objectivation (on serait même plutôt dans le cas inverse), et d'autre part, cette caractéristique est liée au fait d'une utilisation massive de verbes de prise en charge de la subjectivité (ce que semble confirmer la présence importante de verbes comme *penser* ou *croire* (tableau 3). Cette présence importante de verbes est à mettre en parallèle avec la présence importante d'adverbes, venant modaliser les énoncés (*évidemment*, *vraiment*) mais aussi de pronoms personnels de première personne

(tableau 2). On peut aussi noter un plus grand nombre de pronoms pouvant jouer le rôle d'anaphoriques (comme *il(s)* ou *elle(s)*). Ce nombre élevé de pronoms anaphoriques peut expliquer aussi le moins grand nombre de noms : il se peut que le discours oral favorise ce mécanisme de co-référence discursive là où les manuels préféreront une reprise par la tête du terme ou encore par un autre nom. Mais cela resterait à vérifier.

Un autre élément important est à souligner : traditionnellement, on considère que les termes sont majoritairement des groupes nominaux. Or, on peut voir que le pourcentage des quelques structures de groupes nominaux que nous avons testés (N de dét, adj N, N adj (tableau 1)) n'est pas très différent dans les entretiens et dans les manuels (hormis les Nadj qui sont moins présents dans les entretiens).

	Mécanique	Corinthe	Opto
Je	0	978	0
Nous	86	423	93
On	429 (0,4%)	1854 (2,01%)	1027 (0,94%)
Il(s)	596 (0,56%)	970 (1,06%)	567 (0,52%)
Elle(s)	96	204	93

Table 2 : Nombre de pronoms personnels

Adjectifs

En général, dans les corpus spécialisés, les adjectifs permettent de créer des polytermes (ou termes complexes). C'est le cas pour les trois corpus, nous venons de le voir.

Toutefois, certains adjectifs ne visent pas à créer avec le nom un effet d'objectivation mais au contraire à modaliser les noms (par exemple, un *vrai problème*). On pourrait s'attendre à ce que ce soit le cas dans les entretiens. Or, sans que nous ayons pu en faire le décompte précis, il semble que la plupart des structures N adj renvoient, dans les entretiens, à des termes (et non pas une valeur modalisatrice) : *activité contractuelle*, *branches professionnelles*, *pouvoir réglementaire*.

Plusieurs arguments semblent ainsi plaider pour une présence tout à fait réelle de termes dans le corpus d'entretiens :

- les noms sont certes moins nombreux que dans les manuels mais c'est en partie lié à la présence importante de verbes de modalisation (qui sont quasiment absents des manuels) ;
- les pronoms anaphoriques, plus nombreux dans les entretiens, sont peut-être le signe d'un mode d'anaphorisation plus spécifique des entretiens, ce qui, en conséquence, diminue « artificiellement » le nombre de noms ; ce point devrait être vérifié ;
- les groupes nominaux sont presque aussi présents dans les entretiens que dans les manuels.

Négations, conditionnelles

L'examen de marques non ambiguës de négation comme *ne* ou de condition comme *si* (table 3) permet de mettre au jour une nette différence entre les manuels d'un côté et l'entretien de l'autre. L'entretien est beaucoup plus riche en marques de ce type que les manuels. Cette présence importante va avoir une conséquence sur le fonctionnement des marqueurs.

	Mécanique	Corinthe	Opto
Penser	0	150	5
Croire	0	30	1
Evidement	13	43	2
vraiment	0	78	1
Si	179 (0,17%)	304 (0,33%)	184 (0,17%)
Ne	241 (0,23%)	326 (0,35%)	236 (0,21%)

Table 3 : Présence de marques discursives

Bien que cette étude reste préliminaire, il apparaît que l'entretien s'éloigne du fonctionnement classiquement décrit des textes spécialisés supposés être neutres pour se rapprocher de discours spontanés. Toutefois, ce corpus d'entretien reste riche en termes ce qui n'est pas très étonnant puisque les questions posées portent explicitement sur une pratique professionnelle et que l'intention explicative y est présente. Reste à savoir si ce corpus comporte aussi des contextes définitoires, au sens large, c'est-à-dire des contextes riches en marqueurs de relation.

4.1.2 Marqueurs de relation

Cette partie vise à examiner d'une part la présence d'autre part le fonctionnement des marqueurs de relations dans les trois corpus.

Le tableau 4 rend compte de la présence de structures ou de mots présents dans des marqueurs bien connus en français.

	Mécanique	Corinthe	Opto
Nbre de mots	105 700	92 000	109 500
Nom + être + dét	94	30	189
Définir	36	20	6
Définition	0	8	12
Partie	115	47	61
Composer	19	3	12
Composition	3	13	3
Causer	9	1	6
Cause	7	4	6
appeler	38	95	79
Total	321 (0,3%)	221 (0,24%)	374 (0,34%)

Table 4 : présence de marqueurs de relation

Voici trois exemples du marqueur [N être dét] issus de chacun des trois corpus, montrant bien des contextes définitoires.

Le **multiplexage est une** technique qui permet de coupler plusieurs sources lasers à différentes longueurs d'onde ou à plusieurs états de polarisation (optoélectronique)

Le **flambement est une** sollicitation mécanique de compression provoquant le cintrage du forêt lorsque la poussée axiale est excessive (mécanique)

La **Commission de Régulation de l'Energie est une** autorité administrative indépendante (corinthe).

Ce dénombrement permet de faire apparaître plusieurs points :

- une stabilité quasi parfaite du nombre de marqueurs dans les manuels,
 - une présence relativement importante de « marqueurs » dans l'entretien,
 - une présence nettement plus marquée du verbe *appeler* dans les manuels.
- Cette présence s'explique par l'utilisation importante de *on appelle* dans l'entretien (46/95) alors qu'elle est faible dans les autres corpus (6/38 dans mécanique et 0 dans Opto) qui leur préfèrent la forme passive *appelé(e)*. Nous expliquons ci-dessous le rôle de *on* dans ce corpus d'entretien.

Il semble donc qu'il y ait un sens à mettre en oeuvre la méthode des marqueurs dans le corpus d'entretiens même si elle risque d'être moins productive que dans d'autres types de corpus. La partie suivante va montrer qu'outre le fait d'une moins grande présence de marqueurs, un autre problème apparaît du point de vue de la prise en charge des contextes définitoires, au sens large, par les locuteurs dans l'entretien.

4.1.3 Marqueurs et subjectivité

Si l'on regarde, dans le corpus d'entretiens, les énoncés comportant un contexte définitoire, au sens large, on se rend compte que, du point de vue de la présence de marques de prise en charge du discours par le locuteur, une variation importante existe depuis des énoncés où il n'y a aucune marque de prise en charge jusqu'à des énoncés où ces marques sont très présentes. La question de la prise en charge de son discours par le locuteur est bien connue en linguistique dans les travaux sur l'analyse de discours, par exemple dans Kerbrat-Orecchioni (1999) ou dans Maingueneau (1995). Il est donc possible de s'appuyer sur ces travaux pour repérer les marques de subjectivité.

Ainsi dans :

Cac40 c'est quarante firmes quarante sociétés mères.

Il n'y a aucune marque de subjectivité. On peut penser alors que le locuteur reprend cette définition à son compte. Il s'intègre à la communauté des locuteurs collectifs.

D'autres exemples comme :

L'information importante c'est l'information concernant ce que je vais appeler le contrôle amont.

Je vois que tel acteur économique que j'étudie est impliqué quelque part par exemple EDF donc je me dis que c'est une information qui est pertinente qu'il faut intégrer dans la base de données.

La CRE est un acteur particulier qui est un peu en surplomb du secteur avec d'autres autorités hein à mon avis il y a deux ou trois autres autorités qui sont en surplomb.

Ces exemples montrent une position très claire du locuteur par rapport à son discours : présence du pronom de première personne devant des verbes de perception comme *je vois* que ou *je me dis* que, d'adverbiaux subjectifs (*à mon avis*) ou encore de verbes de définition comme *appeler X, Y*.

Certains extraits sont nettement moins clairs quant à l'implication du locuteur dans son discours. Ainsi, certains énoncés se font sur le mode interrogatif :

Est-ce qu'il est faux de dire que la CRE est un acteur du secteur de l'énergie.

De nombreux énoncés utilisent le pronom *on* (table 2) dont (Riegel et al., 2004) a montré qu'il pouvait apparaître à la place de n'importe lequel des 6 pronoms personnels. En l'utilisant, le locuteur peut ainsi jouer sur ce flou et rester dans le vague quant à sa prise en charge de l'énoncé.

Le pouvoir serait représenté complètement si d'une manière ou d'une autre **on** intégrait dans le modèle l'existence du régulateur l'existence de la CRE.

Dans cet extrait, l'effet de vague est renforcé par la conditionnelle, elle-même peu précise.

Enfin, certains énoncés utilisent des figures rhétoriques propres au discours littéraire comme

Vous savez qu' EDF c'est un état dans l'état hein c'est quand même une grande entreprise.

Cet exemple comporte un marqueur de méronymie (inclusion : X être dans Y). Mais, dans ce cas, le fait que X soit égal à Y, répétition complètement absente des autres corpus, empêche que le marqueur fonctionne.

Ces quelques exemples montrent que, s'il y a bien des marqueurs de relations dans le corpus d'entretiens, l'interprétation de l'énoncé dans lequel ils apparaissent comme « subjectif » peut être particulièrement difficile et compliquer sérieusement la tâche du constructeur d'une ontologie personnelle. Cette analyse linguistique devra donc être complétée par une analyse « conceptuelle » essayant de reconstituer une cohérence et surtout, faire l'objet d'une validation par l'auteur même du corpus personnel.

5 Conclusion

La construction d'une ontologie personnelle à partir d'un corpus semble possible avec des méthodes assez proches de celles utilisées pour la construction d'ontologies régionales. Toutefois, des caractéristiques apparaissent, qui rendent la construction d'ontologies personnelles plus difficile. D'une part, du fait de la contradiction qu'il peut y avoir entre « ontologie » et « personnelle », les textes pertinents pour ce type de tâches semblent assez difficiles à identifier. Cependant l'expression directe de son

point de vue par un expert semble pouvoir convenir pour cela. En effet, ces textes comportent des termes (mais en moins grand nombre que les manuels) et ils comportent aussi des marqueurs de relation. La plus grande difficulté tient au fait que la position du locuteur par rapport aux contextes définitoires est parfois très difficile à clarifier. De ce fait, le départ entre ce qui relève d'une ontologie régionale et ce qui relève d'une ontologie personnelle peut être assez difficile à faire. La constitution d'ontologie personnelle apparaît ainsi comme un nouvel objet de collaboration possible entre terminologie textuelle et ingénierie des connaissances. Cet article ayant servi, nous l'espérons, à en jeter les premières bases.

Références

- AIME X., FURST F., KUNTZ P., & F. TRICHET (2009). Gradients de prototypicalité appliqués à la personnalisation d'ontologies. In Actes des 20èmes journées francophones d'Ingénierie des Connaissances (IC'2009), p. 241-252, Presses Universitaires de Grenoble. ISBN 9782706115387. Plateforme AFIA'2009 (Hammamet).
- BACHIMONT B. (1995). Ontologie régionale et terminologie : Quelques remarques méthodologiques et critiques. *La Banque des mots* n°7, Numéro spécial, p. 67-86.
- BOWKER L. (1997). Multidimensional Classification of Concepts and Terms. In S.E. Wright, G.Budin (eds): *Handbook of Terminology management*. Amsterdam/Philadelphia : John Benjamins. p. 131-143.
- CONDAMINES A., (2007). L'interprétation en sémantique de corpus : le cas de la construction de terminologies. *Revue Française de Linguistique Appliquée : Corpus : état des lieux et perspectives*. Vol.XII-1. p. 39-52.
- CONDAMINES A. & D.BOURIGAULT (1999). Alternance nom/verbe : explorations en corpus spécialisés. In B.Victorri et J.François (eds) : *Sémantique du lexique verbal, Actes de l'atelier de Caen, 22-23 janvier 1999*, Cahiers de l'Elsap, p.41-48.
- CONDAMINES A. & REBEYROLLE J., (1996). Point de vue en langue spécialisée », *META*, 42, 1. 1996, p.174-184.
- CONDAMINES A. & REBEYROLLE J., (2000). Construction d'une base de connaissances terminologiques à partir de textes : expérimentation et définition d'une méthode. In J. Charlet, M. Zacklad, G. Kassel & D. Bourigault, (eds). : *Ingénierie des Connaissances, évolutions récentes et nouveaux défis*. Paris : Eyrolles. p. 225-242.
- EUZENAT J. (2009). Edition coopérative de bases de connaissance sur le World-wide web. Actes d'IC 2009.
- FALQUET G. NERIMA L., & J.-C ZISWILER (2004). Utilisation des ontologies dans les environnements d'écriture collaborative d'hyperlivres, expériences et perspectives. *Revue STICEF*. Volume 11, 2004.
- GUSDORF G. (1952). *La parole*. Paris. PUF, 1952.
- JONES P. (2005). Personal Ontology.
http://www.gapingvoid.com/Moveable_Type/archives/001725.html
- JUNG J.J. (2008). Ontology-based context synchronization for ad hoc social collaborations. *Knowledge-Based Systems*, vol.21, issue 7. p. 573-580.
- KERBRAT-ORECCHIONI (1999). *L'énonciation : de la subjectivité dans le langage*. Paris : A. Colin.
- KIM S., HALL W. & A. KEANE (2001). Using document structures for personal ontologies and user modelling. In, *Proceedings of User Modeling 2001: 8th*

- International Conference, UM 2001*. London, UK, Springer, 240-242. (Lecture Notes in Computer Science 2109). <http://eprints.soton.ac.uk/21884/>
- MAINGUENEAU D. (ed.) (1995). Les analyses du discours en France. *Langages* n°117.
- OLSON E.T. (2007). What are we ? A study in personal Ontology. New York: Oxford University press.
- RASTIER F., (1995). « Le terme : Entre ontologie et Linguistique ». *La Banque des Mots* n°7, Numéro spécial. p. 35-64.
- RIEGEL, X, PELLAT J.-C., RIOUL R. (2004). Grammaire méthodique du français. Paris : PUF.